探索扩散模型：从理论到应用的全面综述

刘怡然[1]

1. 北京化工大学信息科学与技术学院，北京 100029

摘 要：扩散模型是一种强大的生成模型，能够在图像、文本和音频等多个领域内产生高质量的结果。本综述旨在汇总和分析应用于视觉领域的扩散生成模型的最新研究进展，包括该领域的理论和实践贡献。本文首先探讨了去噪扩散概率模型、基于分数的扩散生成模型和随机微分方程的扩散生成模型这三种主流模型的特点和原理，并分析了旨在优化模型内部算法和提高采样效率的相关衍生模型。其次，综合评述了扩散模型在当前的应用情况，包括在计算机视觉、自然语言处理、时间序列分析、多模态研究以及跨学科等多个领域的实际应用。最后，基于当前的研究趋势和挑战，对扩散模型未来的发展方向进行了展望，以期为该领域的研究提供指导和启发。本文旨在为研究人员提供一个关于扩散模型研究和应用的全面视图，强调其在人工智能生成内容（AIGC）领域的重要地位和未来潜力。

关键词：深度学习；扩散模型；人工智能生成内容；生成模型；多模态

Exploring Diffusion Models: A Comprehensive Review from Theory to Application

YiRan Liu [1]

1. *College of Information Science and Technology, Beijing University of Chemical Technology, Beijing, 10029*

Abstract: Diffusion models are a powerful type of generative model capable of producing high-quality results in various fields including images, text, and audio. This review aims to summarize and analyze the latest research progress in diffusion models applied in the vision domain, including both theoretical and practical contributions in the field. Initially, the article discusses the characteristics and principles of three mainstream models: denoising diffusion probabilistic models, score-based diffusion generative models, and diffusion generative models based on stochastic differential equations. It also analyzes derivatives aimed at optimizing internal algorithms and improving sampling efficiency. Furthermore, the review provides a comprehensive summary of current applications of diffusion models, including computer vision, natural language processing, time series analysis, multimodal research, and interdisciplinary fields. Finally, based on current trends and challenges, it offers a forecast for the future direction of diffusion models, aiming to guide and inspire research in the field. This article is intended to provide researchers with a comprehensive overview of diffusion model research and application, emphasizing its significant role and potential in the field of Artificial Intelligence Generated Content (AIGC).

Key words： Deep Learning； Diffusion Models； Artificial Intelligence Generated Content； Generative Models； Multimodal Application

0　引　言

在当代科技迅猛发展的背景下，计算机视觉和人工智能成为了推动许多领域前进的关键力量。特别是在生成模型的领域，从基本的模型到现今高度复杂和精细的模型的演化，这些模型不仅在理论上拓宽了我们的视野，也在实际应用中展示了巨大潜力。如图 1 所示，在众多生成模型中，扩散模型以其独特的生成方式和高质量的输出成果而崭露头角，迅速成为学术界和工业界的热点。

随着大数据时代的来临，如何有效地处理和利用海量数据成为了一个挑战。在这个背景下，生成模型特别是扩散模型，展示了处理和生成高质量数据的能力。扩散模型以其高度的灵活性和强大的生成能力，在众多领域中找到了应用，包括但不限于图像生成、超分辨率、图像修复和编辑，以及在自然语言处理和多模态学习中的应用。

此外，随着计算能力的不断增强，扩散模型正逐步克服其计算成本高昂等限制，不断提升其实用性和效率。同时，学界也在积极探索如何进一步提高这些模型的性能和泛化能力，包括通过算法优化和新模型架构的设计。然而，扩散模型的研究和应用还处于发展阶段，许多潜在的应用和改进空间有待挖掘。

因此，本文旨在综述扩散模型的最新研究成果和进展，分析当前的应用情况和面临的挑战，并预测未来的发展趋势。我们将从模型的基础理论入手，逐步深入到各种模型的设计和应用，最终探讨如何将这些理论和技术转化为解决实际问题的有效工具。通过这种方式，本文希望为扩散模型的研究和应用提供一份全面而深入的参考。

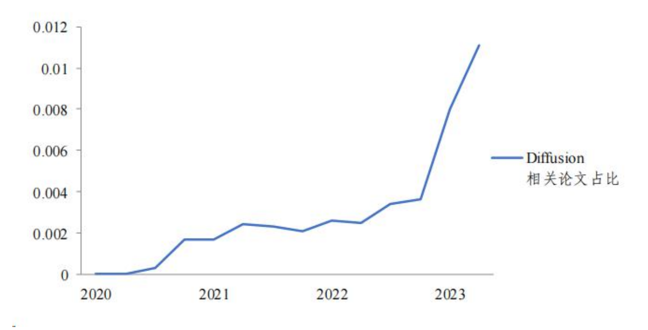


图1 Papers With Code 网站上近年来 Diffusion 相关论文占比变化

Fig. 1 Changes in the percentage of Diffusion-related papers on the Papers With Code website in recent years

1　扩散模型介绍

生成式模型本质上是一组概率分布。如图 2 所示，左边是一个训练数据集，里面所有的数据都是从某个数据 pdata 中独立同分布取出的随机样本。右边就是其生成式模型（概率分布），在这种概率分布中，找出一个分布 pθ 使得它离的 pdata 距离最近。接着在 pθ 上采新的样本，可以获得源源不断的新数据。但是往往 pdata 的形式是非常复杂的，而且图像的维度很高，我们很难遍历整个空间，同时我们能观测到的数据样本也有限。

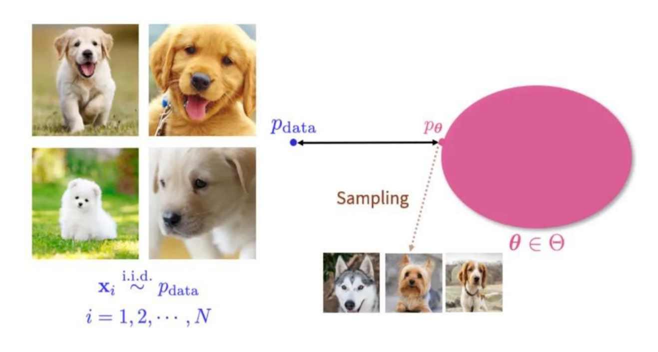


图2 生成式模型流程

Fig. 2 Generative model flow

目前生成式模型目前有四个分支，如图 3 所示，分别是：由 Ian Goodfellow 等人于 2014 年提出的生成对抗网络（Generative Adversarial Models，GAN）[1]，其原理是通过判别器和生成器的互相博弈来让生成器生成足以以假乱真的图像。GAN 已经在图像生成、超分辨率、图像编辑等领域取得了显著的成功。变分自编码器（Variance Auto-Encoder，VAE）[2]是一种基于概率分布的生成模型，由 Kingma 和 Welling 于 2013 年提出。VAE 的核心思想是通过一个编码器将输入图像编码成特征向量，它用来学习高斯分布的均值和方差，而解码器则可以将特征向量转化为生成图像，它侧重于学习生成能力。VAE 在生成可解释性强的样本和生成带有特定属性的样本方面具有广泛的应用。标准化流模型（Normalization Flow，NF）[3,4] 一种通过一系列可逆的转换函数将简单分布转化为复杂分布的生成模型。这些转换函数通过改变概率密度函数的形状，使其逐渐接近目标分布。标准化流模型在生成高维数据和生成具有多模态分布的样本方面表现出色。它的一个关键优势是可以通过逐步变换生成样本，使得生成过程可解释且可控。扩散模型（Diffusion

Models，DM）[5]是一种利用正向过程和反向过程来生成样本的生成模型。在正向过程中，噪声逐渐加入到数据中，而在反向过程中，模型试图逆向预测每一步加入的噪声，从而逐渐还原得到无噪声的样本。扩散模型采用了深度学习的反向传播算法来训练，但其本质上是一个马尔可夫模型。扩散模型的一个关键优势是其生成的样本质量高，且模型理论基础扎实，包括概率模型和随机微分方程等。它们也因可生成高度真实感和多样化的样本而受到青睐。打破了 GAN 在具有挑战性的图像合成任务中的长期主导地位，并且在计算机视觉[6－16]、自然语言处理[17－22]、时间序列[23－25]、多模态[26－33]以及与传统科目[34－42]的结合等领域都展现出不俗的表现。

然而，这些模型通常需要较长的训练时间和大量的计算资源，这是因为反向扩散过程涉及大量的迭代步骤。尽管如此，随着研究的深入，许多方法正在被提出来优化这些模型的效率和性能。综合比较这些模型，扩散模型以其卓越的性能和持续的技术优化，被认为是目前最优秀的生成模型。

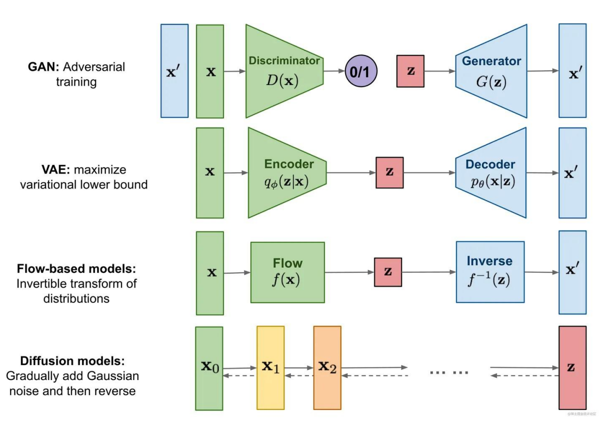


图3 四种主流生成模型框架图

Fig.3 Framework diagram of four mainstream generative models

总的来说，扩散模型作为深度学习领域的一个研究热点，不仅在理论上具有重要意义，同时在实践中也展现出巨大的应用潜力，引领着生成模型的未来发展。

目前扩散模型主要可以划分为三个类别：去噪扩散概率模型（Denoising Diffusion Probabilistic Models, DDPM）[5,43,44]、基于分数的生成模型[45－47]以及基于随机微分方程的生成模型（Stochastic Differential Equations, SDEs）[48]。下面将对这三类模型的构造、理论基础及其在生成过程中的差异性进行深入的讨论和分析。

1.1 去噪扩散概率模型

去噪扩散概率模型是一种深度生成模型，其灵感来自于非平衡热力学，近年来在生成高质量图像、音频和其他复杂数据分布方面展现出了卓越性能。DDPM 的核心思想是模拟数据的扩散过程，如图 4 所示，将结构化的数据逐步转换成无结构的噪声数据，然后通过一个逆过程重新将噪声数据转换回原始数据。

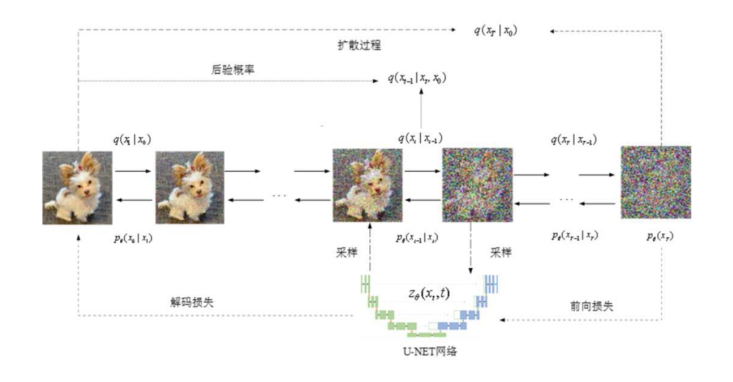


图4 去噪扩散模型处理过程

Fig.4 Process of Denoising Diffusion Probabilistic Models

在前向扩散过程中，模型逐步地向原始数据 $x_0 \sim q(x_0)$ 添加高斯噪声，形成一系列的数据状态 $x_1, x_2, \cdots, x_T$。这一过程可以通过以下马尔可夫链定义：

$$q(x_t \mid x_{t-1}) = N(x_t; \sqrt{1-\beta_t} \cdot x_{t-1}, \beta_t \cdot I), vt \in (1, \ldots, T) \tag{1.1}$$

其中 $t \in \{1, \cdots, T\}$ 代表扩散步骤，β_t 是与每一步相关的方差超参数，I 是单位矩阵，$N(x; \mu, \delta)$ 代表生成 x 的均值 μ 和协方差 δ 的正态分布。$q(x_t \mid x_{t-1})$ 允许我们通过单步操作直接从原始图像采样任意噪声版本 x_t，即 x_t 可以通过原始数据 x_0 和方差计划 β_t 直接采样得到。

在反向生成过程中，DDPM 的核心任务是逐步去除前向过程中引入的噪声，并恢复出清晰的数据。这一过程的目标是训练一个神经网络来模拟从噪声状态 $x_T \sim N(0, I)$ 反向生成原始数据的逆过程。在实践中，这通常通过最小化变分下界进行，其中 Sohl-Dickstein 等人 [49]所提出的负对数似然的变分下界 L_{vlb} 通常包含多个 KL 散度项：

$$L_{vlb} = -\log_{p\theta}(x_0 \mid x_1) + KL(p(x_T \mid x_0) \| \pi(x_T)) + \sum_{t>1} KL(p(x_{t-1} \mid x_t, x0) \| p_\theta(x_{t-1} \mid x_t)) \quad (1.2)$$

其中 KL 表示两个概率分布之间的 Kullback-Leibler 散度。在这个框架下，网络被训练以使 $p_\theta(x_{t-1} \mid x_t)$ 尽可能接近真实的后验分布。

通过这种迭代去噪和重建的过程，DDPM 能够生成与原始数据极其相似的样本，具有非常高的质量和细节保真度。这种方法的成功依赖于精心设计的扩散步骤和高效的网络结构，以及在训练过程中对概率分布进行精确估计的能力。随着技术的不断进步，DDPM 在图像合成、音频生成等领域展示了其强大的潜力，成为深度学习和生成模型领域的一大亮点。

1.2 基于分数的生成扩散模型

基于分数的生成扩散模型，也称为 Score-based Generative Models，是一种先进的深度生成模型，它们通过操纵数据分布的分数函数，即是概率密度函数对数的梯度来生成数据。这些模型的核心思想是，通过逐渐调整添加到数据中的噪声，可以引导数据转换从一个简单的高斯分布逐渐变化成复杂的目标分布。在这些模型中，分数函数表示了在每一点上数据分布的变化速率，为数据的生成路径提供了指引。

在训练阶段，分数模型的目标是学习一个能够准确估计给定数据点分数的神经网络。通常，这个网络被训练来最小化预测分数与真实分数之间的差异，这可以用平方误差来表示：

$$L(\theta) = E_{x\sim pdata}(x), \epsilon \sim N(0, I) \left[\| s_\theta(x + \sigma\epsilon) - \nabla_x \log p_{data}(x + \sigma\epsilon) \|^2 \right] \quad (1.3)$$

其中，E 表示期望，表明是对所有数据点和噪声实例的平均误差，$s\theta$ 是模型学习的分数函数，$\nabla_x \log p_{data}(x + \sigma\epsilon)$ 是数据点 x 处的真实分数，p_{data} 是数据的真实分布，ϵ 代表从标准正态分布 N(0,I) 采样的噪声，而 σ 是噪声的标准差，用于调整噪声强度。

在数据生成过程中，模型从一个简单的高斯分布开始，然后逐步应用逆向扩散步骤来生成数据。这个过程通常被建模为一个连续的随机过程，可以用以下的随机微分方程（SDE）来表示：

$$dx = s(x,t)dt + g(t)dW \quad (1.4)$$

其中，dx 表示数据 x 在微小时间间隔内的变化，$s(x,t)$ 是在时间 t 处的分数函数，表示数据分布在 x 处的概率密度变化的方向和速率。dt 是微小的时间增量。$g(t)$ 是噪声系数，dW 是噪声项。通过解这个 SDE，模型可以逐渐从噪声样本生成目标数据分布的样本。

尽管基于分数的生成模型在生成高质量和多样化样本方面显示出巨大潜力，但它们的计算成本较高，需要大量的数据和时间来训练。此外，设计和优化这些模型的分数函数是一个挑战，因为它要求对数据分布有准确的估计。然而，尽管存在这些挑战，基于分数的生成模型因其生成质量和理论优势，仍然是生成模型领域中的一个重要和活跃的研究方向。随着计算资源的改善和算法的进步，这些挑战有望被逐渐克服，基于分数的模型在未来有望在各种数据生成任务中发挥更重要的作用。

1.3 基于随机微分方程的扩散生成模型

基于随机微分方程（SDEs）的扩散生成模型的发展始于对物理世界中随机扩散过程的研究，这种扩散过程通常使用 SDEs 来数学描述。这些方程在物理学、化学、生物学以及金融学中有着广泛的应用，为描述系统在随机力影响下的时间演化提供了强大的工具。在深度学习领域的早期，这些原理并未被广泛应用，但随着生成模型的需求增长和理论研究的深入，研究者开始探索将这些数学工具应用于数据生成。由于噪声数量的增加会伴随样本生成质量提高，所以 Song 提出了一个可以无穷随机生成噪声的方法，来提高样本的生成质量。

随机微分方程基本形式可以表示为：

$$dx_t = f(x_t, t)dt + g(t)dB_t \quad (1.5)$$

这里，dx_t 是变量 x_t 在时间 t 的微小变化，$f(x_t, t)$ 是决定性的漂移项，描述了系统状态的预期变化趋势，$g(t)$ 是噪声项的强度，而 dB_t 是布朗运动，表示随机的噪声。这个方程描述了一个连续时间下的随机过程，可以用来模拟从简单分布（如高斯分布）到复杂数据分布的过渡。

在深度生成模型中，SDEs 被用来构建一个从目标数据分布逐渐“扩散”到一个简单分布（如高斯噪声）的过程，然后再逆转这个过程来生成新的数据点。在实际应用中，这涉及到精确地模拟和逆转扩散路径，通常要求模型能够学习和近似 SDE 的解。这个过程中，重要的是要确定合适的漂移和扩散项，这通常通过深度神经网络来实现，它们被训练为预测给定时间步的数据分布变化。

这些模型的发展带来了高质量数据生成的新可能，特别是在处理连续和复杂数据分布时展现出了其优势。然而，它们也引入了新的计算挑战，因为精确求解和模拟 SDE 通常是计算密集的。此外，设计有效的数值方法来近似这些连续过程对于实现高效和稳定的生成也至关重要。

随着计算技术的进步和数值方法的发展，基于 SDE 的扩散生成模型正在成为生成模型领域中的一个重要分支。它们不仅为理解和模拟数据分布提供了新的视角，而且为生成高质量、多样化的数据提供了新的工具，预示着在未来的数据生成和模拟应用中将发挥更大的作用。

2 扩散模型的发展及其衍生模型

在扩散模型的研究与应用中，其对于马尔科夫链的依赖性导致了在生成样本时的计算负担，尤其是在处理长步长扩散过程时。随着扩散步长的增加，估计精确的得分函数变得更加复杂和计算密集。这些挑战促使研究者探索各种创新的方法来优化和改进扩散模型的效率和性能。例如，一些研究集中于减少所需的扩散步骤数量，或者开发更高效的数值方法来加速连续时间扩散过程的模拟。其他研究则致力于改善模型架构或训练策略，以提升模型的学习效率和生成质量。

当前的研究工作仍然广泛基于去噪扩散概率模型、基于分数的生成模型，以及基于随机微分方程的生成模型这三种扩散模型的核心理念。这些衍生模型不仅在理论上对原始模型进行了拓展和优化，而且在实际应用中也展示了更广阔的潜力，如图像和音频的高质量生成，以及复杂数据结构的模拟等。

本章节旨在深入探讨扩散模型及其衍生模型的最新研究进展，包括模型架构的创新、训练方法的优化、以及在各领域的应用案例。通过对这些进展的综述，本文将揭示扩散模型在生成模型领域中的重要地位，以及它们如何推动该领域向更高效、更精确的方向发展。同时，将讨论现存的挑战和未来的发展方向，为未来的研究提供方向和启示。通过这种全面而深入的分析，本章节旨在为读者提供一个关于扩散模型及其衍生模型的当前研究现状和未来展望的全景视图。

2.1 基于概率去噪扩散模型的优化

在当代的概率去噪扩散模型优化研究中，模型性能对参数选择的依赖性已引起广泛关注。由于概率去噪扩散模型内部包含多个可调参数，包括噪声水平、采样步骤、以及网络结构等，其性能和效率显著受到参数设置的影响。为了克服这一挑战，研究者们致力于探索和开发各种策略来优化这些关键参数，以提升模型的生成质量和计算效率，因此基于优化各项参数为目标的相关衍生模型也逐渐被提出。

2.1.1 噪声优化

在概率去噪扩散模型的噪声优化领域，近期研究取得了显著的进展，通过精细调控噪声参数和改进噪声注入方式，显著提升了模型的生成性能和效率。Nichol 等人[50]的研究通过在正向加噪过程中引入特定的余弦噪声，优化了模型的对数似然性能，同时在反向去噪过程中引入可学习的方差参数，有效减少了所需的采样步骤，从而提高了整体的采样效率。这种通过精细调控噪声过程来优化模型性能的方法开辟了噪声优化的新方向。

Kingma 等人[51]则探索了将傅里叶特征引入到网络输入以预测噪声的方法，并通过深入分析扩散模型的变分下限（Variational Lower Bound, VLB），揭示了信噪比函数极值对扩散损失的决定性影响。这一发现不仅加深了我们对扩散模型损失结构的理解，也为优化模型提供了重要的理论依据。

此外，动态调整噪声参数[52]的研究也在不断推进。一种新方法使用了 VGG-11 卷积神经网络来训练出最合适的噪声参数，以生成具有更高 FID 值的图像样本，这表明通过优化噪声参数可以直接影响生成样本的质量。此外，面对生成样本过程中可能遇到的对抗性攻击问题，GDMP[53]提纯噪音框架通过在去噪过程中加入净化机制，选择合适的扩散时间步长以淹没对抗性扰动，同时保留输入图像的主要内容，提高了模型在实际应用中的鲁棒性和分类正确率。

2.1.2 改进马尔科夫链

在概率去噪扩散模型的优化领域，改进马尔科夫链的策略显著提升了模型在样本生成速度和质量上的表现。去噪扩散隐式模型（DDIM）[54]的提出是对传统正向马尔科夫过程的一大改进。DDIM 通过实施非马尔科夫过程，即在每一步去噪过程中利用预测的正常样本进行下一步估计，极大地加快了采样速度同时保持了样本的质量。这种改进对于减少模型对大量迭代依赖和提升生成效率具有重要意义。

继 DDIM 之后，更多的研究工作开始集中于进一步改进这些过程。Zhang 等人[55]提出的 gDDIM

模型进一步从数值角度优化了去噪过程。他们发现，在求解相应的随机微分方程时，可以采用特定的分数近似来实现 DDIM，并且指出确定性的抽样方案相较于随机方案能更加迅速地进行采样。这一发现不仅减少了模型的计算负担，也为理解和应用确定性过程在扩散模型中的角色提供了新的视角。

这些研究的共同目标是优化马尔科夫链的设计，以便更快速、更高效地生成高质量的数据。这不仅涉及到算法和模型结构的改进，还包括对模型采样过程和参数设定的精细调整。通过这些改进，模型不仅能够在较短时间内生成样本，还能保证生成样本的多样性和质量，为处理复杂和高维数据提供了强有力的工具。

2.1.3 多模型结合

在多模型结合方面，扩散模型与其他经典生成模型的结合已经成为提升生成性能和拓宽应用范围的重要研究方向。研究者们通过将扩散模型与不同的学习策略和网络结构相结合，旨在提升模型的生成质量、样本多样性以及训练效率。

Sinha 等人[56]提出的具有对比表示学习思想的扩散解码模型是结合了扩散模型和表示学习的一个例子。通过在扩散过程中引入对比自监督学习，该模型不仅能够从扩散先验分布中学习生成样本，还能通过对比学习进一步优化样本的表示质量，显著提高了生成任务的性能，并在多个方面超越了当时的 VAE 模型。

Peebles 等人[57]的研究则将扩散模型与最新的 Transformer 模型结合。他们通过将生成图像任务中常用的 U-Net 网络替换为 Transformer 网络，探索了提高网络深度和宽度、增加 token 数量等策略对于模型性能的影响。这种结合利用了 Transformers 的长程依赖和并行计算优势，进一步提升了扩散模型在处理图像等高维数据时的能力。

此外，GAN 模型与扩散模型的结合[58]也展现了创新的研究方向。在这些结合模型中，扩散模型的稳健性和噪声注入机制被用于改善 GAN 在生成样本稳定性上的不足。通过将扩散模型引入 GAN 的鉴别器中，研究者们能够更好地处理输入数据和生成数据分布的不重叠问题，从而提高了生成样本的稳定性和质量。

2.1.4 针对特殊数据

在处理特殊数据类型上，扩散模型的优化和适应性研究已经取得了显著进展。对于特征主要集中在低密度区域的非常规数据，以及样本量较少的情况，一系列改进的扩散模型被提出以更好地适应这些数据特性，从而在这些具有挑战性的数据分布上生成高质量样本。

Sehwag 等人[59]的研究通过在每个时间步引入两个额外的分类器来优化扩散模型的采样过程，实现了将生成的关注度从高密度区域转向低密度区域，并确保关注度能够停留在低样本密度的数据流形上。这种策略允许模型更有效地在低密度区域生成高质量的样本，提高了在非常规数据上的生成性能。

在少样本数据生成方面，FSDM 框架利用条件 DDPM 进行小规模图像生成，通过结合 VIT[60]框架聚合图像块信息，有效地学习到了已有类别的生成过程，并能够生成更丰富而复杂的样本，以弥补样本量较少的不足。同时，DAG 模型[61]专注于具有几何性质的图像生成，提出了一种利用内部表示进行深度感知图像生成的方法，进一步拓宽了扩散模型在特殊数据类型上的应用。

对于离散数据处理，Austin 等人[62]提出了离散扩散模型。这种模型通过在正向过程中加入多个过渡矩阵，并提出了一种新的损失函数，该损失函数将变分下限与辅助交叉熵损失结合起来，有效地提高了模型在图像生成的对数似然性上的性能，展现了连续扩散模型在离散数据上的适应性和潜力。

2.1.5 超参数优化

在扩散模型的发展过程中，超参数优化已成为提升模型效率和生成质量的重要研究领域。考虑到扩散模型在正反向过程中对于马尔科夫链的依赖可能导致处理效率较低，研究者们致力于通过各种方法来优化模型的采样和训练过程，以实现更快速的样本生成和更精确的模型训练。

Watson 的研究[63]采用了重参数化和重复梯度计算的策略来优化扩散模型的快速采样器，有效减少了模型在采样过程中的计算负担。通过引入 KID 差异指标作为损失函数，并采用随机梯度下降方法进行优化，该方法不仅提升了采样效率，还保持了生成样本的质量。此外，通过特殊的抽样参数化技术，该方法能够显著减少所需的采样步骤，进一步加快了模型的运行速度。

Lam 等人[43]提出的双边去噪扩散模型则从不同的角度对扩散模型进行优化。该模型引入了调度网络和评分网络，对正向和反向过程进行参数化处理。通过精细调控这些网络参数，双边去噪扩散模型能够更有效地学习和模拟数据的生成过程，显著

减少了样本生成所需的步骤，同时提升了生成样本的整体质量。

这些超参数优化方法的提出和应用不仅加深了我们对扩散模型的理解，也推动了模型在实际应用中的性能和效率。通过持续的技术创新和研究，扩散模型在处理效率和生成质量上的表现将持续提升。未来，随着更多高效的超参数优化策略的开发，我们有理由相信扩散模型将在生成模型领域发挥更加关键和广泛的作用，为解决各种复杂数据生成任务提供强大而灵活的工具。

2.1.6 降低 KL 散度

在降低 KL 散度以优化反向去噪过程方面，近期的研究工作已经取得了显著进展。通过深入理解扩散模型中的 KL 散度对模型性能的影响，研究者们致力于开发新的方法来最小化 KL 散度，从而提高模型的推理效率和生成质量。

Watson 等人[64]的研究通过将动态规划算法融入到模型中，实现了对反向去噪过程的优化。他们利用了证据下界（Evidence Lower Bound, ELBO）可以被分解为单独的相对熵项（KL 散度项）的特性，通过最小化这些 KL 散度项来最大化 ELBO。这种方法允许模型在保持生成质量的同时，找到最优的推理路径，大大提升了推理过程的效率和效果。这种对 KL 散度的精细调控，展现了深度理解模型统计性质对优化算法的重要性。

在 Xiao[65]的研究中，通过整合生成对抗网络（GAN）到反向去噪过程中，提出了一种新的方法来最小化 KL 散度。该方法使用 GAN 来区分真实样本和去噪后的样本，通过对抗性训练进一步优化去噪过程，最小化 KL 散度并提高推理效率。这种结合 GAN 的方法不仅提高了推理的准确性，还通过引入对抗性训练提高了模型对于真实数据分布的适应性。

2.1.7 减少采样步骤

在扩散模型的应用中，减少采样步骤以提升生成效率是当前研究的重要方向。由于传统扩散模型需要在整个时间步长中迭代生成数据，这一过程通常耗时且效率低下。因此，研究者们致力于通过各种创新方法来优化时间步长和采样过程，以减少所需的采样步骤并加快模型的生成速度。

Bao[66]的研究通过引入对角和完全协方差来优化时间步长，实现了对 DDPM 生成效率的显著提升。这种优化方法不仅加快了采样过程，也保持了生成样本的质量。通过对时间步长的精确控制和优化，模型能够以更少的步骤生成所需的样本，显著提高了整体的效率。

Chung[67]则从随机微分方程的角度出发，使用了随机差分方程的收缩理论来优化扩散模型的采样步骤。通过对正向过程中初始化图像的优化，研究发现可以显著减少反向去噪过程中的步骤。这种方法通过理论上的深入分析和数学上的严格推导，为减少采样步骤提供了一种有效的途径，进而提高了整体的生成效率。

2.2 基于分数的生成扩散模型优化

2.2.1 改进采样算法

在基于分数模型的生成任务中，改进采样算法是提升模型生成高分辨率且稳定图像的关键。近期的研究工作集中于开发新策略和技术以优化采样过程，提高生成样本的稳定性和质量。

Song 等人[45]的工作代表了这一领域的创新和进步。他们在噪声生成尺度的决定方面采用了新策略，并在采样过程中建议将指数移动平均应用于参数，以提高生成过程的稳定性和连贯性。此外，他们还对分数和损失匹配的加权组合进行最小化处理，以优化分数扩散模型的近似最大似然训练[46]。这些方法不仅提升了模型在生成高质量图像方面的能力，也为理解和优化基于分数模型的采样算法提供了新的视角。

Jolicoeur-Martineau 等人[7]则专注于改进退火采样法，在这个过程中，他们引入了更加稳定的一致性退火采样方案。此外，他们提出了一个由去噪分数和对抗目标组成的混合训练公式，这一公式旨在进一步提高采样的稳定性和效率。通过这种混合训练方法，模型能够在保证生成质量的同时，实现更加高效和稳定的训练过程。

2.2.2 训练梯度优化

在分数生成模型的训练领域，梯度优化是提高模型效率和加速推断过程的关键。由于分数生成模型通常涉及多次迭代的顺序计算，传统方法可能导致推断过程缓慢，因此研究者们开发了新的策略和技术以优化训练过程中的梯度计算。

LSGM[68]提出了一种创新的可变自动编码器框架，旨在潜在空间中训练分数生成模型。该方法的核心是将分数生成模型应用于非连续数据，并在更小的空间中学习更平滑的模型。通过在较低维度的空间中进行训练，LSGM 能够减少网络评估次数，并实现更快的采样过程。这种方法不仅提高了训练和推断的效率，还通过学习平滑的分数函数，提高

了生成样本的质量。

预条件扩散采样（PDS）[69]模型则从另一个角度优化梯度计算。PDS 通过矩阵预处理重新表述扩散过程，有效避免了传统扩散过程中存在的病态曲率问题。这一改进不仅保持了目标分布的质量，还显著提升了模型在实际应用中的效率和稳定性。通过对扩散过程的数学表述进行深入分析和优化，PDS 为训练梯度优化提供了一种有效的途径。

2.2.3 其他改进方面

在分数生成模型的研究中，除了采样算法和训练梯度的优化，还有其他多方面的创新尝试来进一步提升模型的效能和适用性。目前，正向过程在很大程度上依赖于人工设计，这限制了模型的灵活性和适应性。为了解决这一问题，研究者们致力于探索新的理论和方法，以更深层次地理解和优化分数生成模型。

Du 等人[70]的研究通过结合黎曼几何和蒙特卡罗方法的理念，对分数生成模型的正向过程进行了深入的分析和改进。他们提出了一个基于正向过程的参数化扩散模型的通用框架，该框架旨在提供更灵活和高效的方式来设计和实现分数生成模型的正向过程。通过在标准数据集上的测试，他们证明了这种新框架的有效性，不仅在提高模型性能方面取得了成果，也在理解和优化正向过程方面提供了新的视角。

这种将高级数学理论和方法应用于分数生成模型的研究方向，为模型的设计和优化提供了新的可能性。通过更深入地理解模型的数学本质和结构，研究者们能够设计出更加精确和高效的模型，这些模型不仅能更好地适应各种复杂的数据生成任务，也能在理论上提供更加丰富的洞见。

2.3 基于随机微分方程的生成扩散模型的优化

2.3.1 多模型结合

在扩散模型的多模型结合领域，研究者们通过引入新的理论和技术，不断拓展和优化基础模型的功能和性能。Zhang 等人[71]的工作基于微分方程，提出了一种将标准化流与随机微分方程（SDE）相结合的建模方法。这种方法通过联合训练正向和反向 SDE 神经网络，并最小化两者之间差异的共同成本函数，有效地模拟了复杂数据分布的生成过程。这一方法的创新之处在于，它利用后向 SDE 扩散过程以高斯分布开始，并以期望的数据分布结束，为高质量数据生成提供了一种有效的路径。

此外，Kim 等人[72]在 SDE 模型的基础上提出了一种非线性扩散模型。这一模型基于线性扩散模式的标准 SDE 模型，通过结合可训练的标准化流和扩散过程，利用流网络在潜在空间中进行线性扩散来学习噪声分布，然后在数据空间上进行非线性扩散。这种方法的创新在于它在提高模型的灵活性和生成能力的同时，保持了模型结构的简洁性和易于训练的特点。

Ho 等人[73]介绍了一个无分类器的引导方法，旨在克服传统模型中使用分类器引导导致采样结果受限于数据分布局部领域的问题。他们的方法基于从贝叶斯规则衍生出来的隐式分类器，只需要一个条件扩散模型和一个无条件扩散模型，就能生成极高保真度的样本。这种无分类器引导方法为生成模型提供了更大的自由度和更广的适用范围，同时保持了生成样本的高质量。

2.3.2 改进采样算法

在当前的扩散模型研究中，改进采样算法是提升模型效率和生成质量的关键。尤其是在数值 SDE 求解器的应用上，传统方法通常需要大量的分数网络评估，这在实际应用中造成了效率低下的问题。因此，研究者们致力于开发新的策略和技术以优化采样过程，提高生成效率和质量。

Jolicoeur-Martineau 等人[74]设计了一个优化后的 SDE 求解器，其创新之处在于具有自适应步长，能够逐个为基于分数的生成模型量身定制，且只需要两次评分函数评估。这种优化求解器大大降低了计算负担，提高了生成过程的效率，同时保持了生成样本的高质量。

Bortoli 等人[75]针对前向生成过程中噪声分布转换为高斯分布所需大量时间的问题，通过解决路径空间上的熵正则化最优传输问题（也称为薛定谔桥问题）来提高生成效率。这种方法通过优化路径分布的转换，实现了更快的前向生成过程，同时保持了生成样本的分布质量。

Dockhorn 等人[76]将扩散模型与统计力学相联系，提出了一种新的临界阻尼 Langevin 扩散模型（CLD）。该模型通过在数据中添加一个需要学习的速度变量，并学习给定数据的速度条件分布函数，从而简化了直接学习数据分数的复杂度，并且更易于生成高分辨率图像。

Liu 等人[77]则从流形中微分方程求解的角度来看待扩散模型过程。他们发现使用常规数值方法求解反向过程所返回的样本质量较低，而伪数值方法效率很快。为了提升样本质量，他们将数值方法分

为梯度部分和传递部分，旨在使传递部分尽可能地接近目标流形，从而提高了生成样本的质量和效率。

2.3.3 其他改进方面

在扩散模型的研究中，针对高维数据和模型解耦的改进方面有了显著进展，这些创新不仅提升了模型在特定条件下的生成质量，也增加了模型的灵活性和可控性。

Deasy 等人[78]针对高维数据的挑战，提出了噪声引入高斯去噪分数匹配方法以实现扩散强度的可控性。通过引入重尾分布，该方法改进了分数估计和采样收敛，显著提升了无条件不平衡数据集的生成性能。这一方法在提升原随机微分方程（SDE）模型在高维数据生成上的表现的同时，为扩散模型在处理更复杂数据结构提供了新的可能性。

Karras 等人[79]的研究则聚焦于解决扩散模型步骤单元之间的黑盒问题，提出将模型分解成相互独立单元的策略。这种方法增加了模型的可解释性和灵活性，因为对单个单元的修改不会影响其他单元的状态。在此基础上，Karras 主要做出了两个贡献：一是使用 Heun 方法作为常微分方程求解器的采样过程，提升了采样过程的准确性和效率；二是通过对神经网络的输入及其对应标签进行预处理，以训练基于分数的模型，这一策略提升了模型的学习效率和生成样本的质量。

3 扩散模型的应用

3.1 计算机视觉

3.1.1 提高图像分辨率

在单图像超分辨率（SISR）领域，扩散生成模型被广泛研究以解决过度平滑、模式崩溃和高内存占用等问题，并提高生成图像的分辨率。研究者们通过引入新的理念和方法，不断提升超分辨率图像生成的质量和效率。

SRDiff[80]模型利用扩散生成模型和马尔科夫链的特性，将高分辨率图像转换为潜在的简单分布，然后在反向过程中逐步生成超高分辨率图像。这一过程中，低分辨率图像信息作为条件噪声被编码并用于去噪处理，从而有效地提升了超分辨率图像的质量。

SR3[13]模型采用迭代细化策略来提高图像分辨率，解决了单程化的缺陷。它结合了 DDPM 模型的随机去噪过程，并通过训练不同噪声水平的 U-Net 模型来实现迭代优化的去噪过程，从而有效实现超高分辨率图像的生成。这种迭代细化策略不仅提高了生成图像的分辨率，也优化了整体的生成质量和效率。

CDM[81]模型则采用级联的方式将多个扩散模型组成一条流水线。这种级联方式在不同的空间分辨率上采用不同的生成模型，其中包括用于生成低分辨率数据的基础扩散模型，以及用于将图像提高到超高分辨率的 SR3 模型。通过这种多级联的策略，CDM 模型不仅提高了图像的分辨率，还在各个分辨率层次上优化了图像的细节和质量。

3.1.2 图像合成领域

在图像合成领域，扩散模型正在逐渐成为一种重要的替代方法，特别是在解决 GAN 模型训练不稳定和数据覆盖不全等问题上。研究者们正通过结合扩散模型的特性和优势来创新图像合成的方法和框架。

UNIT-DDPM[82]模型是将扩散模型应用于图像合成的一个例子，特别是在非配对的图到图任务中。该模型结合了 DDPM 模型，并引入了元数据域与目标数据域，通过将其中一个域的去噪分数匹配最小化来形成联合分布。然后，利用这种联合分布进行马尔科夫链更新，并最终通过马尔科夫链蒙特卡洛方法生成去噪后的最终样本。这种方法克服了 GAN 在图像合成中的一些局限性，并为生成更加丰富和多样化的图像样本提供了一种有效途径。

Wang 等人[83]的研究将 DDPM 模型应用在语义图像合成领域。他们的模型将噪声图像提供给 U-Net 结构的编码器，而语义布局则通过多层空间自适应归一化算子提供给解码器。此外，通过引入无分类器引导的采样策略，进一步提高了采样质量以及语义可解释性。这种方法有效地提升了语义图像合成的性能和灵活性。

受到自然语言领域 BART[84]模型的启发，ImageBART[14]模型通过学习反转多项式扩散过程来解决自回归图像合成问题。该模型通过引入情景信息，减轻了自回归模型的曝光误差，并解决了自由形式的图像修复问题，而无需特定掩模训练。这种方法在提高自回归图像合成效率的同时，也为复杂和自由形式的图像合成任务提供了新的解决方案。

3.1.3 多维图像领域

在 3D 图像生成领域，扩散模型的应用正在快速发展，为高保真 3D 形状合成、点云处理及场景尺度类别分布学习等方面提供了新的解决方案。研究者们通过引入创新的方法和理论，不断提升 3D

图像生成的质量和效率。

Zhou 等人[85]提出的形状生成补全统一框架（PVD）能够合成高保真形状，补全部分点云，并从真实物体的单视角深度扫描中生成多个完成结果。该框架通过结合扩散生成模型的特性，实现了对复杂 3D 形状的高效生成和补全，为处理和理解 3D 形状提供了新的工具。

Luo 等人[86]则提出了一个用于点云生成的概率模型。他们将点云的生成视为学习将噪声分布转换为所需形状分布的反向扩散过程。这一模型不仅可以应用于点云形状补全、上采样和合成，还可以用于数据增强，如图 5 所示，大大扩展了点云数据的应用范围和效率。这种将反向扩散过程应用于点云生成的方法，为 3D 数据处理提供了新的视角和可能性。

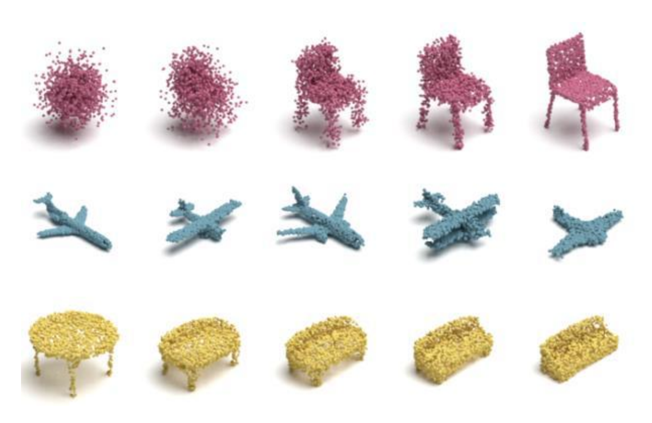


图5 3D模型噪声扩散过程

Fig.5 3D model noise diffusion process

Lee 等人[100]使用离散扩散模型来学习场景尺度类别分布，并使用得出的类别分布来表示场景，从而将多个对象分配到对应的语义类别中。这种方法不仅提高了场景理解和分类的准确性，也为同时生成多个 3D 图像提供了一种有效的方法。通过这种离散扩散模型，研究者们能够更好地理解和生成复杂的 3D 场景，为 3D 图像生成和处理提供了新的途径。

3.2 自然语言处理

扩散模型在计算机视觉领域的广泛应用激发了自然语言处理（NLP）领域研究者的兴趣，他们开始探索将去噪扩散模型应用于文本处理的可能性。然而，与图像的连续空间不同，文本序列具有离散的特性，这给直接应用扩散模型带来了挑战。为了克服这一难题，研究者们提出了以下两种主要的解决思路。

3.2.1 将离散文本映射到连续的表征空间

在将离散文本映射到连续表征空间的研究领域，Difformer[22], DiffusionLM[23]和 DiffuSeq[24]等模型代表了当前的技术进步和理论探索。这些模型通过引入创新的结构和策略，提升了扩散模型在处理离散文本数据上的能力，为文本生成和处理提供了新的方法和视角。

Difformer[22]模型结合了 Transformer 架构和扩散模型的特点，通过引入额外的锚点损失函数、归一化模块以及高斯噪声因子，有效地将离散数据转化为连续数据进行训练。这种结合提升了模型在文本处理上的灵活性和生成能力，同时也保持了 Transformer 架构的强大表达和理解能力。

DiffusionLM[23]模型提出了一种新的基于连续扩散的非自回归语言模型，它将高斯噪声向量迭代去噪为单词向量，创建了向量之间层次连续的潜在关系。这种方法不仅提高了文本生成的连贯性和质量，也在理论上为理解和优化文本生成过程提供了新的途径。

DiffuSeq[24]模型则通过添加一个 embedding 层，将离散文本映射到连续的表征空间。在反向过程中，模型被训练来寻找近似的文本分布序列，从而有效地生成高质量的文本样本。这种方法在保持文本数据的丰富性和多样性的同时，提升了文本生成的效率和质量。

3.2.2 泛化扩散模型

在扩散模型的泛化研究中，相比于传统的将离散文本映射至连续空间的方法，DiffusER[25]模型提出了一种新颖的思路，着重于直接在原始文本上泛化扩散过程。该模型的核心在于将文本的编辑操作，如删除、添加或修改，视作加噪过程，构建了一个更贴近文本数据本质的扩散模型。在反向去噪建模过程中，DiffusER 着力于学习文本的逆变换过程，从而实现目标文本的高效生成，如图 6 所示。这种方法不仅保持了文本数据的离散特性，也提供了一种直观且有效的文本生成方法。

另一方面，DiffusionBERT[26]模型则结合了流行的 BERT[88]模型和扩散模型的优势，在训练过程中提出了一种新的时间步长调度方案。这种方案通过根据每个 token 的信息来控制每一步加噪的程度，从而实现更精细的噪声控制和更有效的学习过程。通过这种结合 BERT 的方法，DiffusionBERT 不仅提升了文本生成的质量，也在模型的理解和表达能力上进行了显著的提升。

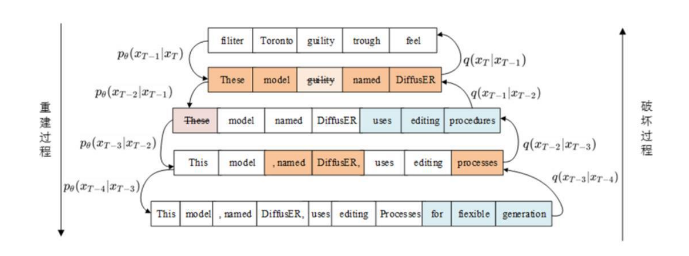


图6 DiffusER 文本生成过程

Fig.6 DiffusER’s text generation process

3.3 时间序列

在时间序列分析领域，扩散模型的应用正日益成为一种新兴趋势，旨在解决传统自回归模型在处理复杂依赖、数据缺失和长期预测方面的限制。扩散模型通过引入新的结构和策略，提升了时间序列分析的精度和效率。

CSDI[23]模型采用基于条件分数的扩散模型替换传统的自回归模型来学习条件分布。该模型将观察数据作为扩散模型的条件输入，利用观察值中的信息进行去噪处理。此外，CSDI 在训练过程中采用自监督方法，将观察值分离为条件信息和插补目标，从而弥补真值缺失的情况。这种方法在处理时间序列数据缺失和预测方面表现出了显著的优势。

SSSD[24]模型则集成了条件扩散模型和结构化状态空间模型，善于捕捉时间序列中的长期依赖关系。该模型在时间序列归并和预测任务中都展现了良好的性能，特别是在处理复杂和长期依赖的数据结构时，显示出其模型的优越性。

TimeGrad[25]模型基于能量生成模型，结合了 RNN[89]和扩散模型的优势来捕获时间序列的特征。在此过程中，它通过优化数据似然的变分界来学习梯度，并在推理时使用 Langevin 采样通过马尔科夫链将白噪声转换为感兴趣分布的样本。这种方法在多元概率时间序列预测方面表现出了优异的性能，尤其是在长期预测和复杂数据结构的学习上。

3.4 多模态

3.4.1 文本转图像

在文本转图像领域，扩散模型已成为推动技术发展的关键因素之一，为生成描述性文本对应的图像提供了新的可能性。扩散模型的应用在提升图像质量、解决生成偏差以及提高生成效率方面展现出了显著优势。

VQ-Diffusion[26]模型在文本转图像的任务中解决了先前生成模型存在的单项偏差问题。该模型利用掩蔽机制来避免在推理过程中误差的累积，从而提高了生成图像的质量和准确性。这种方法不仅提升了生成图像与输入文本的相关性，也提高了图像的细节和质量。

DALLE-2[27]模型则采用了一种两阶段的方法来实现文本转图像的任务,如图 7 所示。在第一阶段，使用 CLIP[90]模型将图像和文本转化为条件嵌入的先验模型；在第二阶段，基于扩散模型的解码器完成图像嵌入工作，从而生成最终的图像。这种两阶段方法充分利用了扩散模型在图像生成方面的优势，同时也保证了文本信息的有效利用。

Imagen[28]模型则由一个用于文本序列的编码器和一个用于生成高分辨率图像的级联扩散模型组成。通过改进原有的 U-Net 模型来进行效率的提升，Imagen 模型不仅提升了生成图像的分辨率，还提高了整体生成过程的效率和质量。

在文本到 3D 图像生成领域，OpenAI 提出的 Point-E 模型代表了该领域的一次重要技术突破，该模型综合运用了两个扩散模型的策略来实现从文本描述到 3D 图像的生成。这一方法标志着文本到 3D 图像生成技术的新发展方向，提供了一种高效和精确的生成策略。





图7 通过 Imagen 生成“舞者在月亮上跳舞”图片

Fig.7 Image generated by Imagen of a dancer dancing on the moon

3.4.2 文本转语音

在文本转语音（Text-to-Speech, TTS）领域，扩散模型被应用于创新的文本到语音生成方法中，旨在提升语音合成的质量、效率和自然度。研究者们通过引入新的架构和策略，改进了传统 TTS 系统的性能，为语音合成技术提供了新的可能性。

Grad-TTS[30]模型提出了一种新颖的文本到语音解决方案，即带有分数的解码器。这种模型逐渐转换编码器预测的噪声，并通过单调对齐搜索生成与文本输入对齐的梅尔频谱图。这种方法有效地将文本信息转化为高质量的语音，提高了语音合成的自然度和准确性。

DiffTTS[32]模型则解决了由于双射约束对模型

宽度限制导致的有效容量不足问题，通过用噪声增量填充中间表示，有效地提升了模型的容量和性能。这种方法不仅提高了语音合成的效率，也提升了生成语音的质量和自然度。

ResGrad[33]模型作为一种轻量级扩散模型，使用残差作为生成目标来改进原来需要从头到尾合成语音的过程。它将现有的 TTS 模型的推理过程变为即插即用的方式，极大地提高了语音合成的灵活性和适应性。这种方法在提升生成语音的质量的同时，也降低了模型的复杂度和计算需求。

3.4.3 文本转视频

在文本转视频生成领域，扩散模型的应用为视频编辑和生成提供了新的视角和方法。随着技术的不断发展，扩散模型在处理更复杂的多媒体任务中展现出其独特的优势和潜力。

Dreamix[91]模型利用扩散模型的特性，在推理阶段根据所提供的文本信息，将低分辨率信息与高分辨率信息相结合进行视频编辑。该模型通过微调模型的初步阶段，有效提高了编辑视频的保真度和准确性。这种方法不仅提升了视频编辑的质量，也为基于文本的视频内容创作提供了一种高效和灵活的解决方案。

Tune-A-Video[92]模型将文本生成视频问题视为生成一系列连续图像的问题，通过提出一种稀疏因果注意力机制，将原本用于图像生成中的空间自注意力扩展到时空域中。这种方法有效地完成了视频的生成工作，提高了生成视频的连续性和自然度，同时也增强了模型对视频内容和结构的理解能力。

3.5 跨学科领域

3.5.1 医学图像领域

医学图像领域的应用中，扩散模型为图像重建、缺陷检测等任务提供了新的解决方案，表现出其在处理高维医学数据和复杂医学问题中的潜力和优势。

针对从测量数据重建图像的逆问题，研究者[38,39]利用分数生成模型作为一种先进的图像重建工具，通过与先验数据一致性来重建图像。这种方法利用扩散模型的特性，能够生成与实际医学图像相符的高质量图像，为医学图像重建提供了一种有效的策略。

Kim 等人[34]提出的由扩散模块和变形模块组成的 DDM 模型，专门用于学习源体积和目标体积之间的空间变形信息。该模型通过生成变化过程的图像来生成 4D（3D 图像加时间）的心脏数据，有效地提升了心脏数据的生成质量和精度，为心脏疾病的诊断和治疗提供了重要的图像支持。

在医学缺陷检测任务上，DDPM 模型被提出来替代传统的自编码器模型[35-37]进行健康图像的训练。在推理时，通过将原始图像中减去生成的健康图像样本来检测异常，从而实现高精度的医学缺陷检测。这种方法不仅提升了缺陷检测的准确性，也为医学诊断提供了一种更为有效和可靠的技术手段。

3.5.2 分子建模领域

在分子建模领域，特别是蛋白质分子的建模，扩散模型被应用于学习和生成蛋白质的动态结构信息，这些研究为蛋白质结构预测和设计提供了新的策略和工具。

Anand 等人[40]的研究利用扩散生成模型去学习蛋白质的旋转和平移等动态的结构信息，从而生成蛋白质的基础结构与序列。这种方法通过捕捉蛋白质的动态特性，可以生成更准确和生物学上可行的蛋白质结构，为蛋白质工程和药物设计提供了重要的结构信息。

ProteinSGM[41]模型将蛋白质的建模过程表述为图像修复问题，并基于条件扩散生成方法对蛋白质结构进行精确建模。这种方法通过类比图像修复的策略，为复杂的蛋白质结构提供了一种新的精确建模方式，提高了蛋白质结构预测的精度和效率。

DiffFolding[42]模型则将蛋白质骨架结构看做一系列连续的角度，用来捕捉组成氨基酸残基的相对方向。结合扩散生成模型，该模型从随机未折叠的结构生成新的稳定折叠结构。这种方法不仅能够提供蛋白质的可能折叠结构，还能够揭示蛋白质折叠过程中的细节和规律，为蛋白质功能研究和药物开发提供了新的视角。

4 存在的问题以及对未来研究方向的展望

基于当前研究，扩散模型已在多个生成领域显示出其显著优势与广阔潜力。然而，其发展仍面临一些关键挑战与问题亟待解决。

首先，扩散模型在正向过程中转化原始图像至全高斯噪声图的目标导致推理过程涉及多个采样步骤和较长的采样时间，从而增加了时间成本。研究如何在预期时间内优化前向加噪过程的停止条件，收敛至特定先验分布，并加入自适应机制是当前研

究的关键。

其次，扩散模型的生成过程依赖长马尔科夫链，使得整个过程呈现出一种黑盒特性，这限制了对依赖关系的捕捉和模型优化的理解。当前，研究需要关注如何将扩散模型分解为独立单元以便进行白盒化处理，以及如何优化默认的马尔科夫链，采用易于捕捉和训练的替代模型。

第三，衍生自扩散模型的改进模型仍多基于 DDPM 的原始设定。未来研究可以考虑将扩散模型作为一种广义模型类型，基于采样算法、扩散方案及构建先验分布等核心思想进行独立研究，使其与其他现有模型更容易结合，扩展应用范围。

第四，当前扩散模型生成样本的评估主要基于 FID 分数，但这一评估无法全面反映样本的恢复效果和多样性。因此，开发新的评估指标以更全面评价模型生成样本质量是未来研究的重要方向。

最后，扩散模型训练通常采用证据下界（ELBO）最小化后验分布与先验分布间的 KL 散度。然而，ELBO 与 NLL 的同时优化理论尚未得到证实，导致实际样本与目标样本存在潜在不匹配问题。此问题关系到模型的实际可靠性和实用性，亟需深入研究和解决。

总体来看，扩散模型未来的研究将聚焦于优化采样算法、降低模型复杂度、提高采样效率等方面。具体可以考虑转化传统的逐步采样算法为更有效的方法，如哈密顿蒙特卡罗方法（HMC），引入预训练模型初始化参数，和采用更优超参数加速训练过程。这些优化方向都是值得进一步探索和研究的。

参考文献(References)

[1] Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial networks[J]. Communications of the ACM, 2020, 63(11): 139–144

[2] Kingma D P, Welling M. Auto-Encoding Variational Bayes[J]. arXiv: 1312.6114,2022.

[3] Zhang M, Sun Y, McDonagh S, et al. Flow Based Models For Manifold Data[J]. arXiv: 2109.14216, 2023.

[4] Rezende D, Mohamed S. Variational Inference with Normalizing Flows[C]// Proceedings of the 32nd International Conference on Machine Learning. 2015: 1530– 1538.

[5] Ho J, Jain A, Abbeel P. Denoising Diffusion Probabilistic Models[C]//Advances in Neural Information Processing Systems. 2020: 6840–6851.

[6] Cheng S-I, Chen Y-J, Chiu W-C, et al. Adaptively-Realistic Image Generation From Stroke and Sketch With Diffusion Model[C]// 2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). 2023:4043-4051.

[7] Jolicoeur-Martineau A, Piché- Taillefer R, Combes R T des, et al. Adversarial score matching and improved sampling for image generation[J]. arXiv: 2009. 05475, 2020.

[8] Chen T, Zhang R, Hinton G. Analog Bits: Generating Discrete Data using Diffusion Models with Self-Conditioning[J]. arXiv 2208. 04202, 2022.

[9] Gu Z, Chen H, Xu Z, et al. DiffusionInst: Diffusion Model for Instance Segmentation[J]. arXiv :2212. 02773, 2022.

[10] Xu J, Wang X, Cheng W, et al. Dream3D: Zero-Shot Text-to-3D Synthesis Using 3D Shape Prior and Text-to-Image Diffusion Models[J]. arXiv :2212. 14704, 2022.

[11] Ye M, Wu L, Liu Q. First Hitting Diffusion Models for Generating Manifold, Graph and Categorical Data[J]. arXiv 2209.01170, 2022.

[12] Furusawa C, Kitaoka S, Li M, et al. Generative Probabilistic Image Colorization[J]. arXiv: 2109.14518, 2021.

[13] Saharia C, Ho J, Chan W, et al. Image Super-Resolution Via Iterative Refinement[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022,45(4):4713-4726.

[14] Esser P, Rombach R, Blattmann A, et al. ImageBART: Bidirectional Context with Multinomial Diffusion for Autoregressive Image Synthesis[C]// Advances in Neural Information Processing Systems. 2021:3518– 3532.

[15] Batzolis G, Stanczuk J, Schönlieb C-B, et al. Non-Uniform Diffusion Models[J]. arXiv: 2207.09786, 2022.

[16] Lee S, Chung H, Kim J, et al. Progressive Deblurring of Diffusion Models for Coarse-to-Fine Image Synthesis[J]. arXiv: 2207.11192, 2022.

[17] Gao Z, Guo J, Tan X, et al. Difformer: Empowering Diffusion Models on the Embedding

Space for Text Generation[J].arXiv 2212.09412, 2023.
[18] Li X L, Thickstun J, Gulrajani I, et al. Diffusion -LM Improves Controllable Text Generation [J]. arXiv: 2205.14217, 2022.
[19] Gong S, Li M, Feng J, et al. DiffuSeq: Sequence to Sequence Text Generation with Diffusion Models[J]. arXiv: 2210.08933, 2022.
[20] Reid M, Hellendoorn V J, Neubig G. DiffusER: Discrete Diffusion via Edit-based Reconstruction[J]. arXiv: 2210.16886, 2022.
[21] He Z, Sun T, Wang K, et al. DiffusionBERT: Improving Generative Masked Language Models with Diffusion Models[J]. arXiv: 2211.15029, 2022.
[22] Lin Z, Gong Y, Shen Y, et al. GENIE: Large Scale Pre-training for Text Generation with Diffusion Model[J]. arXiv :2212.11685, 2022.
[23] Tashiro Y, Song J, Song Y, et al. CSDI: Conditional Score-based Diffusion Models for Probabilistic Time Series Imputation[C]//Advances in Neural Information Processing Systems. 2021: 24804–24816.
[24] Alcaraz J M L, Strodthoff N. Diffusion-based Time Series Imputation and Forecasting with Structured State Space Models[J]. arXiv :2208.09399, 2022.
[25] Rasul K, Seward C, Schuster I, et al. Autoregressive Denoising Diffusion Models for Multivariate Probabilistic Time Series Forecasting [C]//Proceedings of the 38th International Conference on Machine Learning. 2021: 8857–8868.
[26] Gu S, Chen D, Bao J, et al. Vector Quantized Diffusion Model for Text-to-Image Synthesis[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).2022:10686-10696.
[27] Ramesh A, Dhariwal P, Nichol A, et al. Hierarchical Text-Conditional Image Generation with CLIP Latents[J]. arXiv : 2204.06125, 2022.
[28] Saharia C, Chan W, Saxena S, et al. Photorealistic Text-to-Image Diffusion Models with Deep Language Understanding[J]. arXiv : 2205.11487, 2022.
[29] Nichol A, Jun H, Dhariwal P, et al. Point-E: A System for Generating 3D Point Clouds from Complex Prompts[J]. arXiv :2212. 08751, 2022.
[30] Popov V, Vovk I, Gogoryan V, et al. Grad-TTS: A Diffusion Probabilistic Model for Text-to-Speech [C]//Proceedings of the 38th International Conference on Machine Learning.2021: 8599–8608.
[31] Jeong M, Kim H, Kim H, et al. Diff-TTS: A Denoising Diffusion Model for Text-to-Speech[J]. arXiv: Audio and Speech Processing, 2021.
[32] Chen Z, Wu Y, Leng Y, et al. ResGrad: Residual Denoising Diffusion Probabilistic Models for Text to Speech[J]. arXiv:2212. 14518, 2022.
[33] Kim B, Ye J C. Diffusion Deformable Model for 4D Temporal Medical Image Generation[J]. arXiv : 2206.13295, 2022.
[34] Wolleb J, Bieder F, Sandkühler R, et al. Diffusion Models for Medical Anomaly Detection[J]. arXiv: 2203.04306, 2022.
[35] Sanchez P, Kascenas A, Liu X, el al. What is Healthy? Generative Counterfactual Diffusion for Lesion Localization[C]//Deep Generative Models. Cham: Springer Nature Switzerland.2022: 34–44.
[36] Wyatt J, Leach A, Schmon S M, et al. AnoDDPM: Anomaly Detection with Denoising Diffusion Probabilistic Models using Simplex Noise[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW).2022: 649–655.
[37] Song Y, Shen L, Xing L, et al. Solving Inverse Problems in Medical Imaging with Score-Based Generative Models[J]. arXiv: 2111.08005, 2021.
[38] Chung H, Ye J C. Score-based diffusion models for accelerated MRI[J]. arXiv :2110.05243, 2022.
[39] Anand N, Achim T. Protein Structure and Sequence Generation with Equivariant Denoising Diffusion Probabilistic Models[J]. arXiv :2205.15019, 2022.
[41] Lee J S, Kim P M. ProteinSGM: Score-based generative modeling for de novo protein design[J]. bioRxiv, 2022.
[42] Wu K E, Yang K K, Berg R van den, et al. Protein structure generation via folding diffusion[J]. arXiv : 2209.15611, 2022.
[43] Lam M W Y, Lam M W Y, Wang J, et al. Bilateral Denoising Diffusion Models.[J]. arXiv :2108.11514, 2021.

[44] Giannone G, Nielsen D, Winther O. Few-Shot Diffusion Models[J]. arXiv : 2205.15463, 2022.
[45] Song Y, Ermon S. Improved Techniques for Training Score-Based Generative Models[C]// Advances in Neural Information Processing Systems. 2020: 12438– 12448.
[46] S Song Y, Durkan C, Murray I, et al. Maximum Likelihood Training of Score-Based Diffusion Models [C]//Advances in Neural Information Processing Systems.2021:1415– 1428.
[47] Song Y, Ermon S. Generative Modeling by Estimating Gradients of the Data Distribution [C]//Advances in Neural Information Processing Systems.2019.
[48] Song Y, Sohl-Dickstein J, Kingma D P, et al. Score-Based Generative Modeling through Stochastic Differential Equations[J]. arXiv :2011.13456, 2021.
[49] Sohl-Dickstein J, Weiss E, Maheswaranathan N, et al. Deep Unsupervised Learning using None-quilibrium Thermodynamics[C]// Proceedings of the 32nd International Conference on Machine Learning. PMLR.2015:2256–2265.
[50] Nichol A Q, Dhariwal P. Improved Denoising Diffusion Probabilistic Models[C] //Proceedings of the 38th International Conference on Machine Learning.2021:8162– 8171.
[51] Kingma D, Salimans T, Poole B, et al. Variational Diffusion Models[C]//Advances in Neural Information Processing Systems.2021: 21696–21707.
[52] San-Roman R, Nachmani E, Wolf L. Noise Estimation for Generative Diffusion Models[J]. arXiv : 2104.02600, 2021.
[53] Wang J, Lyu Z, Lin D, et al. Guided Diffusion Model for Adversarial Purification[J]. arXiv :2205. 14969, 2022.
[54] Song J, Meng C, Ermon S. Denoising Diffusion Implicit Models[J]. arXiv: 2010. 02502, 2020.
[55] Zhang Q, Tao M, Chen Y. gDDIM: Generalized denoising diffusion implicit models[J]. arXiv :2206. 05564, 2022.
[56] Sinha A, Song J, Meng C, et al. D2C: Diffusion-Denoising Models for Few-shot Conditional Generation.[J]. arXiv: 2106.06819, 2021.
[57] Peebles W, Xie S. Scalable Diffusion Models with Transformers[J]. arXiv :2212. 09748, 2022.
[58] Wang Z, Zheng H, He P, et al. Diffusion-GAN: Training GANs with Diffusion [J]. arXiv :2206.02262, 2022.
[59] Sehwag V, Hazirbas C, Gordo A, et al. Generating High Fidelity Data From Low-Density Regions Using Diffusion Models[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2022:11492–11501.
[60] Dosovitskiy A, Beyer L, Kolesnikov A, et al. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale[J]. arXiv :2010.11929, 2021.
[61] Kim G, Jang W, Lee G, et al. DAG: Depth-Aware Guidance with Denoising Diffusion Probabilistic Models[J]. arXiv :2212. 08861, 2023.
[62] Austin J, Johnson D D, Ho J, et al. Structured Denoising Diffusion Models in Discrete State-Spaces [C]//Advances in Neural Information Processing Systems.2021:17981– 17993.
[63] Watson D, Chan W, Ho J, et al. Learning Fast Samplers for Diffusion Models by Differentiating Through Sample Quality[J].
arXiv:2202.05830, 2022.
[64] Watson D, Ho J, Norouzi M, et al. Learning to efficiently sample from diffusion probabilistic models[J]. arXiv:2106.03802, 2021.
[65] Xiao Z, Kreis K, Vahdat A. Tackling the generative learning trilemma with denoising diffusion GANs[J]. arXiv:2112.07804, 2021.
[66] Bao F, Li C, Sun J, et al. Estimating the Optimal Covariance with Imperfect Mean in Diffusion Probabilistic Models[J]. arXiv : 2212.08861, 2022.
[67] Chung H, Sim B, Ye J C. Come-Closer-Diffuse-Faster: Accelerating Conditional Diffusion Models for Inverse Problems Through Stochastic Contraction[C] //2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).2022: 12413–12422.
[68] Vahdat A, Kreis K, Kautz J. Score-based Generative Modeling in Latent Space[C]//Advances in Neural Information Processing Systems.2021: 11287–11302.
[69] Zhang L, Zhu X, Feng J. Accelerating Score-based Generative Models with Preconditioned Diffusion Sampling[J]. arXiv : 2207.02196, 2022.

[70] Du W, Yang T, Zhang H, et al. A Flexible Diffusion Model[J]. arXiv :2206. 10365, 2022.
[71] Zhang Q, Chen Y. Diffusion Normalizing Flow[J]. arXiv :2110.07579, 2021.
[72] Kim D, Na B, Kwon S J, et al. Maximum Likelihood Training of Implicit Nonlinear Diffusion Models[J]. arXiv :2205. 13699, 2022.
[73] Ho J, Salimans T. Classifier-Free Diffusion Guidance[J]. arXiv :2207.12598, 2022.
[74] Jolicoeur-Martineau A, Li K, et al. Gotta Go Fast When Generating Data with Score-Based Models.[J]. arXiv :2105.14080, 2021.
[75] Bortoli V D, Thornton J, Heng J, et al. Diffusion Schrödinger Bridge with Applications to Score-Based Generative Modeling.[J]. arXiv: 2106.01357, 2021.
[76] Dockhorn T, Vahdat A, Kreis K. Score-Based Generative Modeling with Critically-Damped Langevin Diffusion[J]. arXiv :2112.07068, 2022.
[77] Liu L, Ren Y, Lin Z, et al. Pseudo Numerical Methods for Diffusion Models on Manifolds[J]. arXiv: 2202.09778, 2022.
[78] Deasy J, Simidjievski N, Liò P. Heavy-tailed denoising score matching[J]. arXiv :2112.09788, 2022.
[79] Karras T, Aittala M, Aila T, et al. Elucidating the Design Space of Diffusion -Based Generative Models[J]. arXiv :2206. 00364, 2022.
[80] Li H, Yifan Y, Chang M, et al. SRDiff: Single Image Super-Resolution with Diffusion Probabilistic Models.[J]. arXiv :2104.14951, 2021.
[81] Ho J, Ho J, Saharia C, et al. Cascaded Diffusion Models for High Fidelity Image Generation[J]. arXiv :2106.15282, 2021.
[82] Sasaki H, Willcocks C G, Breckon T P. UNIT-DDPM: UNpaired Image Translation with Denoising Diffusion Probabilistic Models. [J]. arXiv :2104.05358, 2021.
[83] Wang W, Bao J, Zhou W, et al. Semantic Image Synthesis via Diffusion Models [J]. arXiv :2207.00050, 2022.
[84] Lewis M, Liu Y, Goyal N, et al. BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension[C]// Proceed- ings of the 58th Annual Meeting of the Association for Computational Linguistics. 2020: 7871–7880.
[85] Zhou L, Zhou L, Zhou L, et al. 3D Shape Generation and Completion through Point-Voxel Diffusion[J]. arXiv: :2104.03670, 2021.
[86] Luo S, Luo S, Luo S, et al. Diffusion Probabilistic Models for 3D Point Cloud Generation[J]. arXiv: 2103.01458, 2021.
[87] Lee J, Im W, Lee S, et al. Diffusion Probabilistic Models for Scene-Scale 3D Categorical Data[J]. arXiv :2301.00527, 2023.
[88] Devlin J, Chang M-W, Lee K, et al. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding[C]// Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Minneapolis, Minnesota: Association for Computational Linguistics. 2019:4171 –4186.
[89] Schmidt R M. Recurrent Neural Networks (RNNs): A gentle Introduction and Overview[J]. arXiv:1912.05911, 2019.
[90] Radford A, Kim J W, Hallacy C, et al. Learning Transferable Visual Models From Natural Language Supervision[J].arXiv : 2103.00020, 2021.
[91] Molad, Eyal,et al. Dreamix: Video Diffusion Models are General Video Editors [J]. arXiv: 2302. 01329, 2023.
[92] Wu, Jay Zhangjie, Ge, et al. Tune-A-Video: One-Shot Tuning of Image Diffusion Models for Text-to-Video Generation [J]. arXiv :2212.11565, 2022.